**Topic 1: Attributes, Variables, Levels of Measurement, Populations and Samples**

**June 5, 2011.**

**Attributes and Variables:** De Veaux in his introduction to attributes and variables states that the data used in statistical analysis have two important elements, values *and their context* (p. 3). It is the context that surrounds the value which is essential to determining whether a variable is quantitative or quantitative and is the first step toward demonstrating sound statistical reasoning. One important question to answer about data is the “who” or “what” that is being examined. For example, persons, objects and situations canfunction as *cases.* Each case, in turn, has *characteristics or attributes that are known as variables* that allow us to distinguish differences or variation among them.

De Veaux states that there are two main categories of variables. *Quantitative* variables are those that can be measured *in units* and are possessed *in degree.* In contrast *qualitative* or *categorical* variables answer questions about *categories,* or what kind, and are typically represented as counts or frequencies, not as units on a scale. These distinctions are important because not all numerical values can be assumed to be quantitative variables. Some examples include ID numbers, dates, or age, as well as frequency counts of categorical variables. Kinnear notes that perhaps most confusing of all is the use of the term “categorical” to describe data because it blurs the differences between nominal and ordinal data as well as the distinctions between ranks and ratings (p. 3).

**Levels of Measurement:** De Veau’s advice about considering the actual purpose of a variable is useful when examining what has been written about levels of measurement. This appears to be an area of considerable debate based on looking at a number of different sources. Vogt defines levels of measurement as the term used to compare measurement scales based on *how much information different categories of variables convey*. Vogt distinguishes four types of variables (ratio, interval, ordinal and nominal); Kinnear outlines three (scale, ordinal and nominal).

Kinnear states that *scale or continuous* data use an independent scale with units (p. 2). Vogt notes that continuous variables can be expressed as an infinite number of measures (p. 69). De Veaux does not address this type of variable at all in his initial chapters. Kinnear notes that scale variables not only express the degree to which a property or attribute applies to a case, but also convey considerable information independently of other measures in the same set. For example, *ratios* combine two numbers that show their relative size. These are typically conveyed as fractions or decimals.

*Interval* measures are often considered scale variables because they use an independent scale with equally-spaced units. The distinction is that although the distance between any two adjacent units of measurement is the same, there is no meaningful zero point. A Fahrenheit scale is a useful example, because zero degrees does not mean the absence of heat or cold, but expresses a temperature located between -1 and +1 degrees.

*Ordinal* data are the third level of measurement. They are expressed in the form of rankings based on *perceived* weight of attributes. Unlike interval or ratio variables, ordinal measures are not based on an independent scale with units. Rankings only have meaning relative to other data in the set.

The last category is *nominal* data which are simply labels that record category membership. They *do not measure the degree to which an attribute is present.* Researchers often assign arbitrary numbers to for different categories or labels, such as 1= Associates degree; 2= Bachelors degree, etc.

**Populations and Samples:** In quantitative research, a *population* is the group to which a researcher wishes to generalize his or her results. This is sometimes referred to as the theoretical population. In contrast the population that is actually available to the researcher to sample is known as the *accessible population*. Parameters are the key numbers used in a model for a population.

In contrast, a sample is defined as a subset of a population. If the statistics that are computed from a sample of a larger population accurately reflect the corresponding parameters, is said to be *representative* of the theoretical population.

**Questions, concerns, observations**

1. Is there a difference between scale and continuous variables? Are these really synonyms?

2. As I have begun to play with my data set in SPSS, I realize that it is important to double check the level of measurement (scale, ordinal, or nominal) assigned to each variable before working with the dataset. Many of the demographic variables were listed as scale, when in fact they were actually nominal.

3. I am struggling to figure out how to classify the level of measurement of the data gathered from the Likert scales in survey questions 12-25. When I look at my data set, these were all automatically assigned as scale variables. My initial inclination was to consider them ordinal data.

Yet Kinnear raises the question whether ratings can be considered scale data. Kinnear notes that some say that the use of reference or anchor points, found at the far ends of the scale, provide a form of reference for raters’ judgments, so they can be considered interval data. He also notes that others consider the individual rankings within the scale as merely ordered categories, so that 100 respondents are merely generating 100 sets of ranks with ties (multiple counts in some ranks?) , i.e., the ratings themselves are merely ordinal data (p. 2). Vogt notes that ordinal data are often treated as continuous variables when there are many ranks in the data, and as ordinal when there a just a few (p. 61).

Based on my data set, I do not think we can assume that the intervals between the 5 response categories are equal, although they do appear to have a rank order. The inclusion of the “I choose not to answer this question” option in the scale also complicates things, because this is not ranking, per se. The best advice I saw came from Jamison who suggests that decisions regarding levels of measurement should be considered at the design stage and must be addressed in the methodology section of their research (p. 1218).

4. Preparing data for analysis. One stipulation for using this data was that I maintain the confidentiality of the participating institutions by eliminating any indentifying information that was specific to a course, instructor, program, or SUNY campus. I have removed the student comments field in the database and have started to recode the campuses into categories: community colleges, comprehensive colleges, technical colleges, specialty colleges and university centers.

5. What are my options for dealing with the Likert response “I do not wish to answer this question.” This appears everywhere in the survey. If I delete these responses, each survey question will have a different denominator. Is this a problem? Or is this a matter of documenting this as a footnote?

**Works Cited**

DeVeaux, R., Velleman, P., & Block, D. (2009). *Intro Stats*, 3rd ed., Boston: Pearson Education.

Jamison, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1212-1218.

Kinnear, P.R.& Gray, C. D. (2010*). IBM SPSS Statistics 18 Made Simple*. New York: Psychology Press.

Vogt, W. P. (2005). *Dictionary of Statistics and Methodology: A Non-Technical Guide for the Social Sciences.* Thousand Oaks, CA: Sage.